

Zensus

─ Nutzung des Zensus 2011 zur Modellierung der Wohnstandortwahl von Haushalten unter Geheimhaltungsbedingungen

von **Benjamin Heldt** und **Dirk Heinrichs** (Deutsches Zentrum für Luft- und Raumfahrt – DLR)

Der Zensus 2011 ermöglicht die Untersuchung der unterschiedlichsten Fragestellungen von der Betrachtung von Pendlerverflechtungen (Kutzki, 2015) bis zur Analyse der Lebens- und Wohnsituation älterer Menschen (Deecke und Niemann-Ahrendt, 2015). Der Mikrodatensatz des Zensus bietet aufgrund seines großen Umfangs aber auch ein bisher kaum genutztes Potenzial für komplexe raumbezogene Untersuchungen. Im Rahmen des Projektes „Verkehrsentwicklung und Umwelt“ des Deutschen Zentrums für Luft- und Raumfahrt beispielsweise sollen auf Grundlage des Zensus 2011 Aussagen über Entscheidungen zur Wohnstandortwahl von Haushalten in Berlin getroffen werden. Die Ergebnisse nutzen die Autoren im Standortwahlmodell „SALSA“ zur Untersuchung der Interaktion zwischen Verkehr und Landnutzung. Die eingeschränkte Zugänglichkeit von Mikrodatensätzen unter Geheimhaltungsbedingungen erfordert die einfache Ausführbarkeit von Aufbereitung und Analyse verbunden mit Ergebnissen, die so ausgegeben werden sollten, dass eine Prüfung einfach und schnell durchführbar ist. Gleichzeitig müssen die Ergebnisse ausreichend genau sein, um die Modelle evaluieren und weiterverwenden zu können. Ein entsprechendes Analysekonzept entwickelte und testete das Institut für Verkehrsforschung im DLR im Rahmen eines Pilotprojektes mit dem Amt für Statistik Berlin-Brandenburg (AfS). Im Folgenden werden der Ansatz für die Datenaufbereitung des Zensus 2011 für die Parameterschätzung von Standortwahlkriterien unter Berücksichtigung der Datenschutz- und Modellierungsanforderungen und Beispiele für die vorbereitende Analyse anhand von deskriptiven Statistiken vorgestellt.

Anforderungen der Schätzung eines Standortwahlmodells

Alltägliche Mobilität und die Standortwahl von Haushalten interagieren miteinander, insbesondere da der Wohnstandort zumeist Ausgangspunkt alltäglicher Wege ist. Um diese Interaktion zu untersuchen, werden am DLR-Institut für Verkehrsforschung Simulationsmodelle aufgebaut, die es ermöglichen, Auswirkungen von Änderungen beim Verkehr oder der Landnutzung auszuwerten. Eine damit verbundene wichtige Fragestellung ist, wie Haushalte ihren Wohnstandort (definiert als Wohnung in einem Wohngebiet) wählen und welche Bedeutung ver-

schiedene Eigenschaften der Standorte (also der Wohnung und des Raumes) in diesem Kontext haben. Insbesondere interessiert dabei die verkehrliche Erreichbarkeit von Zielorten der alltäglichen Wege.

Zumeist wird hierfür der Ansatz der diskreten Entscheidungswahl unter Nutzenmaximierung verwendet (McFadden, 1978). Demnach wählt ein Haushalt aus einem zufällig gezogenen Set an Standorten denjenigen aus, der seinen Nutzen maximiert (NM-Modell). Für jede Standorteigenschaft kann so die Einflussstärke ermittelt werden. Die Art der Zufallsziehung beeinflusst dabei die Schätzung der Parameter der Nutzenfunktion. Einen alternativen Ansatz stellt das durch Alonso (1964) theoretisch fundierte „bid-rent“-Modell dar. Demnach entscheidet sich ein Standortanbieter für den Haushaltstyp mit der höchsten Zahlungsbereitschaft, gegeben, dass dies den Nutzen des Haushaltes maximiert (ZB-Modell). Anders formuliert: Die Wohnung wählt den Haushalt. Im Entscheidungsmodell sind daher die Wohnungen die Entscheider und Haushaltstypen bilden die Alternativen. Dies ermöglicht genauso wie das nutzenbasierte Modell die Ermittlung von Einflussstärken der Standorteigenschaften – in diesem Fall allerdings auf die Zahlungsbereitschaft eines Haushaltstyps. Das ZB-Modell hat gegenüber dem NB-Modell den Vorteil, dass keine zufallsbasierte Auswahl der Standorte vorgenommen werden muss und es mit einem Angebots- und Mietenmodell zu einem Simulationsmodell für die Standortwahl kombiniert werden kann (Martínez, 1992, 1996).

Modelle der diskreten Entscheidungswahl können formuliert als multinomiale logistische Regression (MNL) mithilfe des Maximum Likelihood-Ansatzes geschätzt werden (Ben-Akiva und Lerman, 1985). MNL haben nicht zum Ziel, absolute Werte zu bestimmen; vielmehr werden Wahrscheinlichkeiten für das Eintreten der Kategorien berechnet. Die Koeffizienten im Modell beschreiben dabei immer relative Unterschiede zu einer Referenzkategorie. Im Falle der Standortwahlmodellierung werden somit Differenzen zwischen standortbezogenen Zahlungsbereitschaften von Haushalten mit unterschiedlichen Eigenschaften ermittelt. So ist beispielsweise davon auszugehen, dass Haushalte mit relativ hohem Einkommen größere Wohnungen in kleineren Gebäuden bevorzugen und dies ihre Zahlungsbereitschaft

daher stärker erhöht – und somit auch die Wahrscheinlichkeit, dass dieser Haushaltstyp im Vergleich zu anderen gewählt wird. Eine andere beispielhafte Fragestellung wäre: „Wie viel höher oder niedriger ist die Zahlungsbereitschaft von Familienhaushalten für einen Quadratmeter Wohnraum mehr im Vergleich zu Einpersonenhaushalten, gegeben, dass die anderen Faktoren konstant bleiben?“ Für die Modell-schätzung müssen mithilfe von aus theoretischen Überlegungen abgeleiteten Annahmen und statistischer Analyse aus einer Vielzahl von Parametern die potenziell einflussstärksten Variablen ausgewählt werden. Die Modellschätzung selbst erfolgt mit geeigneter Software, zum Beispiel „Biogeme“ (Bierlaire, 2003).

Mikrodaten für die Modellierung der Standortnachfrage von Haushalten

Die Modellierung der Standortnachfrage erfordert aufgrund der Komplexität der Entscheidungssituation die Analyse umfangreicher, aus verschiedenen Disziplinen¹ kombinierter Datensätze zur Ableitung von Parametern. Voraussetzung für die Untersuchung und Simulation der Interaktion zwischen Wohnen und Mobilität sind Informationen, die kombinierte Merkmale von Haushalten, deren Mobilitätsausstattung und dazugehörigen Wohnungen sowie Eigenschaften des Wohngebietes enthalten (Heldt, Gade und Heinrichs, 2014). Diese Eigenschaften sind im Simulationsmodell die veränderbaren Parameter, anhand derer Szenarien, wie zum Beispiel die Auswirkung von Bevölkerungswachstum auf die räumliche Verteilung von Haushalten, getestet werden können.

Bisher gibt es keine öffentlichen Datensätze, die alle diese genannten Merkmale enthalten. Teilkombinationen bilden die durch die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder (FDZ) bereitgestellten demografischen Mikrodaten des Mikrozensus 2010 und des Zensus 2011 sowie das System repräsentativer Verkehrserhebungen (SrV) des Jahres 2008 ab.

Diese Datensätze eignen sich unterschiedlich gut für die Parameterschätzung. Da das SrV nur Haushalts- und Mobilitätsmerkmale, aber keine Wohnungseigenschaften enthält und räumlich gering aufgelöst ist, wurde dieser Datensatz für weitere Analysen ausgeschlossen. Der Mikrozensus ist eine einprozentige Stichprobe und lässt damit unterhalb der Bezirksebene in Berlin keine räumlichen Verortungen zu. Entsprechend ist er für die Schätzung der Parameter ebenfalls nicht geeignet, da Raummerkmale auf Bezirksebene nur eine geringe Streuung aufweisen und durch höher aufgelöste Wohnungsmerkmale überdeckt würden. Aufgrund seines umfangreichen Merkmalskranzes ermöglicht der Mikrozensus aber als Hilfsdatensatz die Imputation von Merkmalen, die in anderen Datensätzen nicht enthalten sind. Im Rahmen des Zensus 2011 wurden

die bereinigten Registerdaten in Form einer Haushaltegenerierung mit den Daten der Gebäude- und Wohnungszählung verknüpft. Der verknüpfte Datensatz enthält eine sehr hohe Anzahl an Einträgen, die räumlich genau verortet Merkmalskombinationen von Haushalten und Wohnungen zulässt. Der Zensus 2011 wird daher als zentraler Datensatz für die Schätzung der Verhaltensparameter des Standortwahlmodells herangezogen.

Der Zugang zu den Mikrodaten des Zensus 2011

Die Einzeldaten des Zensus 2011 und des Mikrozensus 2010 unterliegen dem Statistikgeheimnis nach § 16 BStatG² (vgl. Statistische Ämter des Bundes und der Länder (2015)). Infolgedessen werden Analysemethoden benötigt, die mit den Anforderungen der statistischen Geheimhaltung kompatibel und dennoch flexibel genug für die Durchführung von Modellschätzungen sind.

§ 16 BStatG besagt unter anderem, dass die Reidentifikation von Individuen in den Mikrodatensätzen verboten ist. Eine Ausnahme stellt das sogenannte Wissenschaftsprivileg dar, welches in § 16 Abs. 6 BStatG aufgeführt ist. Demnach dürfen Wissenschaftlerinnen und Wissenschaftler am Gastwissenschaftlerarbeitsplatz (GWAP) der FDZ auf faktisch anonymisierte Mikrodatensätze zugreifen. Die erzeugten Ergebnisse werden dann durch Mitarbeiterinnen und Mitarbeiter des AfS auf Einhaltung der statistischen Geheimhaltung geprüft und elektronisch übermittelt. Beim Zensus 2011 wird die statistische Geheimhaltung hauptsächlich durch das Verfahren SAFE (Sichere Anonymisierung für Einzeldaten) sichergestellt. Das Verfahren arbeitet anders als die klassischen Geheimhaltungsmethoden. Bei SAFE wird ein Rückschluss auf Einzelpersonen verhindert, indem die Einzeldaten leicht verändert werden (vgl. Statistische Ämter des Bundes und der Länder (2015) und Höhne (2015)).

Um diese Anforderungen an die statistische Geheimhaltung bei der ersten Nutzung des kombinierten Mikrodatenmaterials des Zensus 2011 einzuhalten, vereinbarte das DLR mit dem AfS die Durchführung eines Pilotprojektes. Die Fachabteilung Zensus stellte die Datensätze auf Ebene der Teilverkehrszellen zur Verfügung.³ Der erzeugte Datensatz besteht dabei aus zwei Teilen: Mikrodaten auf Personenebene und Mikrodaten auf der Ebene der Familien, Haushalte, Wohnungen und Gebäude. Beide Datensätze sind durch eine eindeutige Haushalts-ID verknüpfbar. Die Verfahrensweise für die Auswertung der Daten umfasst zwei Stufen. Am Gastwissenschaftlerarbeitsplatz werden auf dem Sample die Skripte zur Datenaufbereitung und Modellschätzung entwickelt. Anschließend führen die Mitarbeiterinnen und Mitarbeiter des AfS die Analysen auf dem Originalmaterial durch (kontrollierte Datenfernverarbeitung – KDFV). Für deskriptive Statistiken wird das anonymisierte SAFE-Material

¹ in der Regel Bevölkerungszählungen und Erhebungen für die Verkehrsforschung.

² Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I S. 2394).

³ Mehr zur räumlichen Einteilung Berlins in Bömermann (2014) und zur Bereitstellung der Daten des Zensus 2011 in Tag und Voy (2012).

ausgewertet, für Modellschätzungen das Originalmaterial.⁴

Analysekonzept zum Umgang mit den Geheimhaltungs- und Modellierungsanforderungen

Die Analysen müssen nicht nur den Anforderungen hinsichtlich der statistischen Geheimhaltung genügen, sondern auch den Vergleich unterschiedlicher Haushaltstypisierungen zulassen und eine ausreichende Flexibilität zum Testen verschiedenster Modelle ermöglichen. Dabei sollten die Analyseprogramme nutzerfreundlich sein, um die Ausführung per KDFV zu erleichtern und damit die Bearbeitungszeit zu verkürzen. Da für die Modellierung nur Unterschiede zwischen den Gruppen von Bedeutung sind, sind mit den Modellierungsanforderungen gleichzeitig die Anforderungen an die statistische Geheimhaltung erfüllt. Die Modellschätzung in Form der multinomialen logistischen Regression ist daher direkt datenschutzkonform und ermöglicht schon aufgrund des schätzungseigenen Fehlerterms keinen Rückschluss auf Einzelfälle. Für die notwendige Vorauswahl der Parameter können deskriptive Statistiken zum Beschreiben von Verhältnissen und Zusammenhängen genutzt werden, die nicht oder nur indirekt auf Fallzahlen zugreifen. Dies umfasst einerseits hinsichtlich möglicher Reidentifikationen unbedenkliche Korrelationsanalysen. Andererseits können die räumliche Variation der Standortmerkmale und der Gruppen in Standardabweichungs-Karten dargestellt werden – da hier nur Kategorien von Werten veranschaulicht werden, sind hierdurch ebenfalls keine Rückschlüsse auf Einzelwerte möglich.

Programme zur Aufbereitung und Analyse

Für die einfache und schnelle Aufbereitung und Ergebniserzeugung wurde ein Programm geschrieben, das die Softwarepakete IBM SPSS und R verwendet. Den zentralen Bestandteil bilden dabei modular aufgebaute und konfigurierbare SPSS-Syntaxen und R-Skripte zur Datenaufbereitung und -visualisierung.

Der Nutzende passt die Parameter in den Konfigurationsdateien an seine Anforderungen an und führt dann die zentralen Steuerungsskripte aus, die auf die verschiedenen SPSS- beziehungsweise R-Skripte zugreifen. Hierdurch ist die Datenaufbereitung nutzerfreundlich und auch durch die Mitarbeiterinnen und Mitarbeiter des AfS leicht ausführbar.

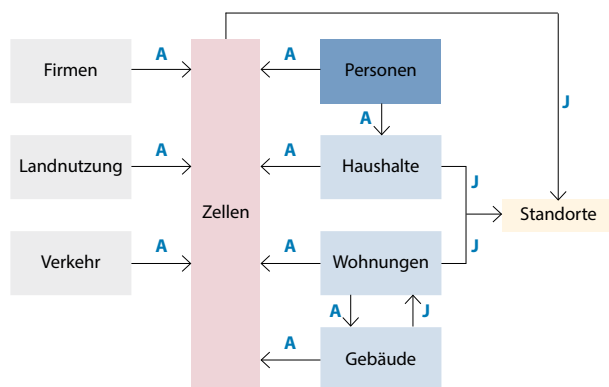
Der Zieldatensatz enthält auf Wohnungs- bzw. Haushaltsebene⁵ Merkmale von Haushalten, Wohnungen und Wohngebieten. Die Erstellung des Datensatzes erfordert eine Abfolge verschiedener Verknüpfungs- und Aggregationsschritte (Abbildung a). Der erste Teil des Programms führt folgende Schritte zur Datenaufbereitung aus:

1. Aggregation der Personenmerkmale auf Haushaltsebene,
2. Berechnung gebäudebezogener Wohnungsmerkmale,
3. Zusammenführung von Haushalts- und Wohnungsmerkmalen.

Die Durchführung der Schritte steuert ein SPSS-Skript. Um auf Haushaltsebene heterogene, nicht auszahlbare Merkmale wie beispielsweise das Alter zu aggregieren, wird in einem ersten Schritt eine den Haushalt repräsentierende Person bestimmt. Für die weiteren Analysen wird davon ausgegangen, dass diese Person am ehesten die Entscheidung über den Wohnstandort trifft. Als Haushaltsrepräsentant gilt die älteste beschäftigte Person im Haushalt bzw. die älteste Person, wenn keine Person im Haushalt beschäftigt ist. Im Rahmen des Zensus 2011 wurden aus den Registern der Bundesagentur für Arbeit die sozialversicherungspflichtig Beschäftigten (ohne geringfügig Beschäftigte) sowie Bedienstete im öffentlichen Sektor, etwa Beamte, Richter und Soldaten, identifiziert. Die restliche arbeitende Bevölkerung liegt in keinem Register vor. Dafür wurde eine 10 %-Stichprobe gezogen, in der diese erfasst und auf die Gesamtbevölkerung hochgerechnet wurde. Der Einfachheit halber definierten die Forscher allerdings für dieses Projekt die entsprechenden Personen als „nicht beschäftigt“. Weitere Haushaltsmerkmale, zum Beispiel die Anzahl der Kinder oder Seniorinnen und Senioren, wurden über die Haushalts-ID ausgezählt.

Ein wichtiges Kriterium bei der Wohnstandortwahl ist die Größe des Wohngebäudes in Form der Wohnungsanzahl. Die Gebäude- und Wohnungszählung (GWZ) differenziert dabei nur wenige Kategorien. Insbesondere für die Identifikation von Gebäuden mit einer hohen Anzahl von Wohnungen ist es daher sinnvoll, diese Variable selbst zu berechnen. Dafür können Wohnungen über Identifikatoren Gebäuden zugeordnet und je Gebäude die Anzahl der Wohnungen bestimmt und zurück an die Wohnungen angespielt werden. Schließlich wurden die Haushalts- und Wohnungsmerkmale über die Haushalts-ID zusammengeführt.

a | Workflow für die Aufbereitung des Zensus 2011 zur Erstellung eines Datensatzes für die Standortwahlschätzung



FDZ-Daten:
Zensusgesamtdatenbestand

PERSON – auf Personenebene
FHWG – auf Haushaltsebene
Externe Daten

J Verknüpfung (Jbin)
A Aggregation

⁴ Mehr zur Bereitstellung der Daten des Zensus 2011 über die FDZ in Raab und Meisdrock (2015).

⁵ Die Haushaltsdefinition im Zensus 2011 geht davon aus, dass alle gemeinsam lebenden Personen in einer Wohnung einen Haushalt bilden und somit – außer in den Sonder- und Gemeinschaftsunterkünften – eine Wohnung einen Haushalt ent-

hält. Dies umfasst beispielsweise auch Wohngemeinschaften, siehe § 2 Gesetz über den registergestützten Zensus im Jahre 2011 (Zensusgesetz 2011 – ZensG 2011) vom 8. Juli 2009 (BGBl. I S. 1781).

Das R-Skript greift auf den durch das SPSS-Skript aufbereiteten Zensusdatensatz zu und führt die folgenden Schritte zur Verknüpfung und Anreicherung der Zensusdaten und zur deskriptiven Analyse durch:

1. Aufbereitung des Mikrozensus und Schätzung des Einkommens-Imputationsmodells,
2. Imputation des Einkommens im Zensus,
3. Segmentierung des Datensatzes,
4. Berechnung und Ergänzung raumbezogener Merkmale,
5. Berechnung deskriptiver Statistiken und Darstellung in Karten,
6. Erzeugen des Schätzdatensatzes.

Das Skript kann auf verschiedenen Anwendungsumgebungen (KDFV, GWAP, eigene Umgebung) eingestellt werden. Es ermöglicht die Bestimmung der Haushaltseinteilung, der Imputationsmethode sowie der Stichprobengröße. Hierdurch können in kurzer Zeit beispielsweise Datensätze für die Modellschätzung der Standortwahl von Haushalten unterschiedlicher Größe, unterschiedlichen Einkommens oder einer Kombination von beidem erstellt werden.

Im ersten Schritt wird die Arbeitsumgebung vorbereitet, R-Pakete installiert und geladen und Funktionen zur Berechnung und Darstellung geladen. Da in den Ergebnissen des Zensus 2011 kein Einkommen enthalten ist, dienen Daten des Mikrozensus 2010 zur Schätzung eines Imputationsmodells (Heldt et al., 2017). Hierfür werden die Mikrozensus-Daten aufbereitet und die Einkommensschätzung in Form eines ordinal-logistischen Regressionsmodells (McCullagh, 1980) für vier Einkommensklassen durchgeführt. Relevant hierbei ist, dass der Haushaltsvorstand und die im Modell verwendeten Variablen exakt so wie im Zensus 2011 definiert sind. Im Anschluss findet die Aufbereitung der Zensusdaten selbst sowie die Imputation des Einkommens statt. Im Ergebnis wird jedem Zensus-Haushalt für jede Einkommensklasse eine Wahrscheinlichkeit zugeordnet.

Im nächsten Schritt erfolgen die Umkodierung weiterer Variablen im Zensusdatensatz und dadurch die Bildung von Haushalts- und Wohnungstypen. Der bis dahin erzeugte Datensatz enthält nun Haushaltsmerkmale (inklusive Einkommen) und Wohnungsmerkmale. Die einzige raumbezogene Information ist allerdings die Nummer des Wohngebiets (Teilverkehrszelle). Auf Basis dieser ID spielt das Skript extern berechnete Merkmale, die das Wohngebiet und seine Erreichbarkeit beschreiben, an. Aus den Zensusdaten selbst werden über Aggregation ebenfalls Raummerkmale, wie zum Beispiel der Anteil der Haushalte nach Haushaltsgröße oder -einkommen, berechnet und angespielt.

Diskrete Wahlmodelle können zusätzlich zu den erklärenden Variablen, die die Entscheider beschreiben, Eigenschaften berücksichtigen, die die Alternativen selbst – in diesem Fall die Haushalte – definieren. Dafür müssen allerdings zufallsbasiert Haushalte gezogen und deren Merkmale an den Datensatz angespielt werden. Dafür enthält das Skript eine Funktion zur Generierung von Alternativen. Da die später mit Biogeme durchgeführte Modellschätzung eine hohe Rechenkapazität erfordert, zieht das Programm für die Erzeugung des finalen Datensatzes eine nutzer-

definierte Stichprobe aus den Gesamtdaten, an die die Alternativen angespielt werden. Das Ergebnis dieses Prozesses ist eine Biogeme-konforme, Tabstopp-getrennte Datei. Für die Einschätzung der Imputationsgüte erfolgt außerdem die Ausgabe des Einkommensmodells.

Deskriptive Statistik zur Vorauswahl von Modellparametern

In der Schätzung des Standortnachfragemodells ist es aufwändig, alle Parameter zu testen. Daher führt das Programm eine vorbereitende Analyse mithilfe deskriptiver Statistiken durch, die konform mit den Anforderungen der statistischen Geheimhaltung sind. Kern der Analyse sind Maße zur Beschreibung der Lage und Streuung der erklärenden und der abhängigen Variablen sowie ihres Zusammenhangs. Beispielhaft sollen anhand eines Beitrags zur mobilTUM 2016 (Heldt, Gade und Heinrichs, 2016) das Zusammenspiel der Analysemethoden und der Datenschutz- und Modellierungsanforderungen verdeutlicht werden.

Als erste wichtige deskriptive Statistik zeigt die Korrelationsmatrix die Zusammenhangsmaße zwischen allen Variablen in Form von Pearson-Korrelationskoeffizienten. Zentral sind hierbei die Korrelationen zwischen der aggregierten Form der abhängigen Variablen (Anteil des Haushaltstyps in der Zone) und allen anderen. Erklärende Merkmale, die hierfür hohe Werte aufweisen, werden in die Modellschätzung einbezogen und auf Signifikanz getestet. Gleichzeitig deckt diese Analyse Korrelationen der Variablen untereinander und somit potenzielle Multikollinearität im Schätzmodell auf. Das Beispiel (Abbildung b) zeigt Korrelationen zwischen Anteilen je Haushaltstyp nach Alter des Haushaltsrepräsentanten (Tabelle 1) und Erreichbarkeitsindikatoren (Tabelle 2).

1 | Haushaltstypen

Code	Beschreibung
A1	Haushalte, deren Repräsentant ¹ jünger als 31 Jahre ist
A2	Haushalte, deren Repräsentant zwischen 31 und 64 Jahren alt ist
A3	Haushalte, deren Repräsentant älter als 64 Jahre ist

¹ Der Haushalt wird durch die älteste erwerbstätige Person bzw. die älteste Person repräsentiert, wenn keine Person erwerbstätig ist.

2 | Erreichbarkeitsindikatoren (jeweils der kürzeste Weg)

Code	Beschreibung
1a	Gewichtete Pkw-Distanz zu anderen Zonen
1b	Gewichtete Fußweg-Distanz zu anderen Zonen
2	Gewichtete Fußweg-Distanz zum nächsten Geschäftszentrum
3a	Gewichtete Pkw-Distanz zum nächsten Lebensmittelgeschäft mit einer Verkaufsfläche von mindestens 1000 m ²
3b	Gewichtete Fußweg-Distanz zum nächsten Lebensmittelgeschäft mit einer Verkaufsfläche von mindestens 200 m ²
4a	Durchschnittliche erreichbare Verkaufsfläche in Lebensmittelmärkten in 10 min gewichteter Pkw-Distanz
4b	Durchschnittliche erreichbare Verkaufsfläche in Lebensmittelmärkten in 10 min gewichteter Fußweg-Distanz
5a	Gewichtete Pkw-Distanz zum nächsten Park mit einer Größe von mindestens 50 000 m ²
5b	Gewichtete Fußweg-Distanz zum nächsten Park mit einer Größe von mindestens 10 000 m ²

b | Tabelle der Korrelationen zwischen Haushaltstypanteilen je Zone und Erreichbarkeitsindikatoren

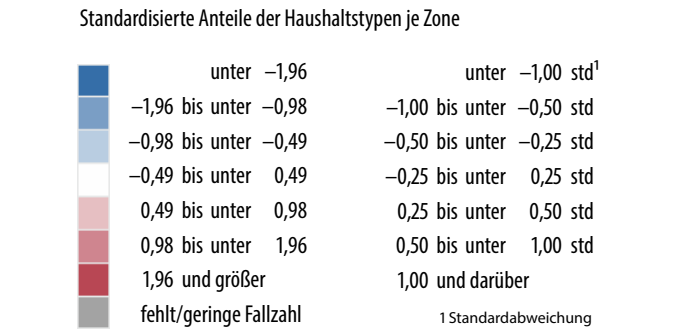
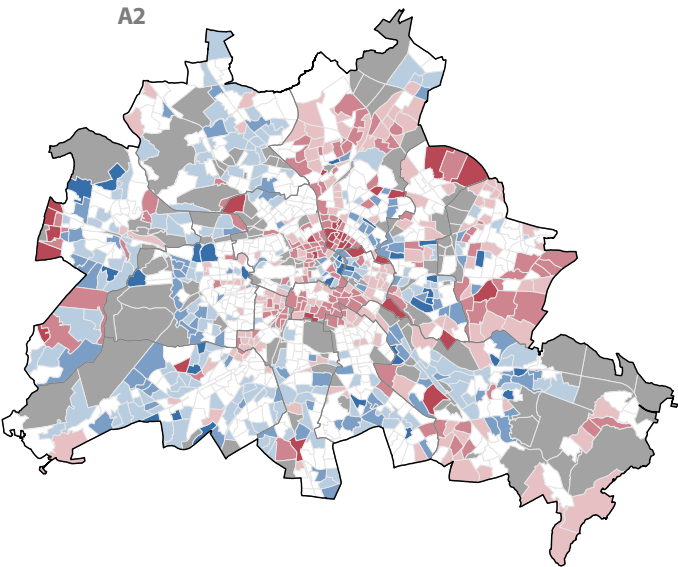
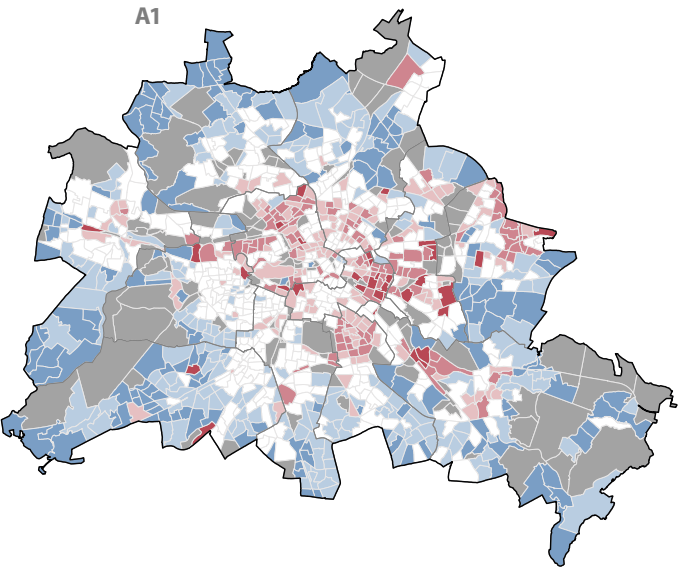
A1	***	***	***	***	***	***	***	***	***	***	*	***
-0,23	A2	***	n.s.	n.s.	*	*	**	n.s.	n.s.	*	**	
-0,67	-0,61	A3	***	***	***	***	***	***	***	n.s.	**	
-0,41	-0,00	0,34	1a	***	***	***	***	***	***	***	***	
-0,44	-0,02	0,38	0,97	1b	***	***	***	***	***	***	***	
-0,27	0,08	0,17	0,48	0,50	2	***	***	***	***	***	***	
-0,36	0,08	0,25	0,51	0,48	0,56	3a	***	*	*	***	***	
-0,39	0,10	0,26	0,52	0,51	0,58	0,69	3b	***	n.s.	***	***	
-0,23	0,02	0,18	0,67	0,60	0,24	0,06	0,23	4a	***	*	**	
-0,12	-0,06	0,15	0,35	0,31	0,19	-0,07	-0,00	0,54	4b	n.s.	n.s.	
-0,08	0,07	0,01	0,27	0,24	0,53	0,49	0,45	0,07	0,00	5a	***	
-0,19	0,09	0,09	0,29	0,31	0,54	0,37	0,47	0,10	0,02	0,61	5b	

Quelle:
Zensus 2011,
OSM contributors,
Senatsverwaltung
für Stadtentwick-
lung und Wohnen
Berlin 2014, eigene
Berechnungen

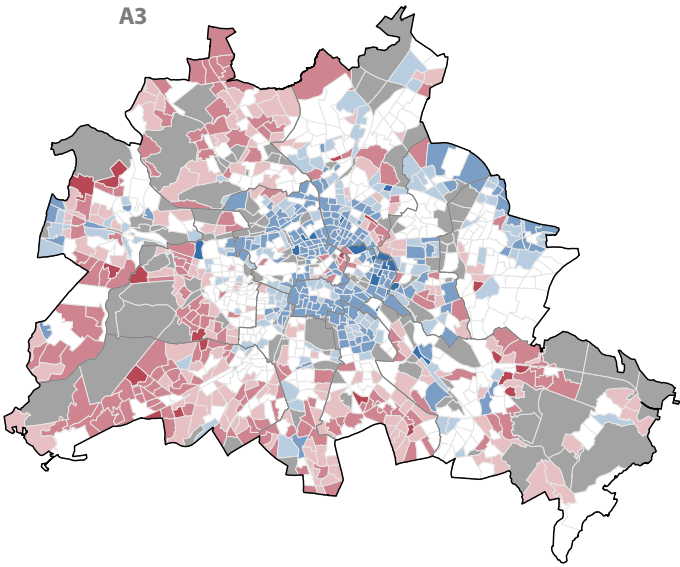
In der oberen Hälfte von Abbildung b sind die Signifikanzniveaus (n.s. = nicht signifikant; * = 0,005 < p < 0,05; ** = 0,0005 < p < 0,005; *** = p < 0,0005) zu erkennen, in der unteren farbig kodiert die Korrelationskoeffizienten. Zu sehen ist, dass für die Haushaltstypen A1 und A3 signifikante Zusammenhänge mit der Mehrzahl der Indikatoren bestehen, nicht jedoch für den Haushaltstyp A2. Weiterhin ist erkennbar, dass mehrere Indikatoren untereinander starke Korrelationen aufweisen. Für eine Analyse der Standortwahl von Haushalten in unterschiedlichen Lebensphasen sollten daher insbesondere die Indikatoren 1a bis 3b einbezogen werden. Für die Erklärung des zweiten Haushaltstyps ist es notwendig, andere Merkmale zu berücksichtigen oder den Typ noch weiter zu unterteilen.

Die zweite deskriptive Statistik, Karten der Verteilung der Haushaltstypenanteile, dient zur grundlegenden Analyse der Modellstruktur, das heißt der Definition der abhängigen Variablen und der Beschreibung ihrer Verteilung (Abbildung c).

c | Räumliche Verteilung der Haushaltstypen A1 bis A3



Quelle: Zensus 2011, eigene Berechnungen



Dargestellt sind die Anteile als positive (rot) bzw. negative (blau) Abweichungen vom Mittelwert für alle Zonen. Die Karten veranschaulichen also Besonderheiten in der räumlichen Verteilung der Haushaltstypen nach Lebensphase. Bis auf wenige Ausnahmen bestätigen sie die Vermutung, dass jüngere Haushalte eher im Zentrum zu finden sind und ältere am Stadtrand. Für die vorliegende Studie zeigen die Karten außerdem, dass die Wahl der Haushaltstypen passend für ein Standortwahlmodell ist, da sie sich hinsichtlich ihrer räumlichen Muster voneinander unterscheiden. Ähnliche Verteilungen könnten darauf hinweisen, dass Typen zusammengeführt werden sollten.

Fazit

Das beschriebene Pilotprojekt zeigt, dass Standortwahlmodelle, das heißt Modelle der raumbezogenen diskreten Entscheidungswahl, anhand von Mikrodaten unter Einhaltung der statistischen Geheimhaltung analysiert werden können. Voraussetzung hierfür ist der Einsatz deskriptiver Statistiken zur Beschreibung von Verteilungen, die keinen Rückschluss auf Einzelwerte ermöglichen, wie zum Beispiel Karten von Standardabweichungen und Korrelationsanalysen.

Für den Einsatz am Gastwissenschaftlerarbeitsplatz der FDZ wurde ein nutzerfreundliches, konfigurierbares Skript entwickelt, das die Ausführung durch Mitarbeiterinnen und Mitarbeiter der FDZ erleichtert. In diesem Beitrag liegt der Schwerpunkt auf der Datenaufbereitung und der vorbereitenden Analyse; Ergebnisse der Modellschätzungen werden demnächst veröffentlicht (Heldt et al., 2017).



Benjamin Heldt ist Diplom-Geograph und arbeitet seit 2011 am Institut für Verkehrsforschung im Deutschen Zentrum für Luft- und Raumfahrt (DLR) in der Abteilung *Mobilität und urbane Entwicklung*. Dort verantwortet er die Umsetzung von Modellen zur Simulation der Standortwahl. Dabei untersucht er statistische Methoden zur Verknüpfung von Datensätzen und interessiert sich besonders für die Interaktionen zwischen Raummerkmalen, Mobilität und den Entscheidungen von Haushalten und Unternehmen. Kontakt: benjamin.heldt@dlr.de



Dirk Heinrichs leitet seit 2014 die Abteilung *Mobilität und urbane Entwicklung* am Institut für Verkehrsforschung im Deutschen Zentrum für Luft- und Raumfahrt (DLR) und ist Professor für das Fachgebiet Stadtentwicklung und Mobilität am Institut für Stadt- und Regionalplanung der TU Berlin. Sein Forschungsinteresse gilt den Wechselbeziehungen zwischen räumlicher Entwicklung und Mobilität sowie den sozialen und ökologischen Folgen in Städten, insbesondere vor dem Hintergrund rapider Urbanisierung in Lateinamerika, Asien und Afrika. Kontakt: dirk.heinrichs@dlr.de

Quellen

- Alonso, W. (1964): Location and land use. Toward a general theory of land rent. Cambridge, United Kingdom: Harvard University Press.
- Ben-Akiva, M.; Lerman, S. R. (1985): Discrete Choice Analysis. Theory and Applications to Travel Demand. Cambridge, MA: The MIT Press.
- Bierlaire, M. (2003): BIOGEME: a free package for the estimation of discrete choice models. In: Proceedings of the 3rd Swiss Transportation Research Conference.
- Bömermann, H. (2014): Berlin kleinräumig. Zeitschrift für amtliche Statistik Berlin Brandenburg, Heft 2/2014, S. 20–22.
- Deecke, J.; Niemann-Ahrendt, K. (2015): Generation 65 plus. Leben und Wohnen älterer Menschen in Berlin und Brandenburg. Zeitschrift für amtliche Statistik Berlin Brandenburg, Heft 3/2015, S. 22–30.
- Heldt, B.; Donoso, P.; Bahamonde-Birke, F.; Heinrichs, D. (2017): Estimating bid-auction models of residential location using census data with imputed household income. Beitrag wird präsentiert auf dem World Symposium on Transport and Land Use Research, Juli 2017, Brisbane, AUS.
- Heldt, B.; Gade, K.; Heinrichs, D. (2014): Challenges of Data Requirements for Modelling Residential Location Choice: the Case of Berlin, Germany. Beitrag veröffentlicht in European Transport Conference, Frankfurt am Main, Deutschland.
- Heldt, B.; Gade, K.; Heinrichs, D. (2016): Determination of Attributes Reflecting Household Preferences in Location Choice Models. In: Transportation Research Procedia, München, Deutschland.
- Höhne, J. (2015): Das Geheimhaltungsverfahren SAFE. Zeitschrift für amtliche Statistik Berlin Brandenburg, Heft 2/2015, S. 16–33.
- Kutzki, V. (2015): Pendlerverflechtungen sozialversicherungspflichtiger Beschäftigter in der Metropolregion Berlin-Brandenburg. Zeitschrift für amtliche Statistik Berlin Brandenburg, Heft 3/2015, S. 32–39.
- Martínez, F. (1992): The bid-choice land-use model: An integrated economic framework. Environment and Planning A, Band 24, S. 871–885.
- Martínez, F. (1996): MUSSA – Land use model for Santiago City. Transportation Research Record: Journal of the Transportation Research Board, Band 1552, Heft 1, S. 126–134.
- McCullagh, P. (1980): Regression models for ordinal data. Journal of the Royal Statistical Society. Series B (Methodological), Band 42, Heft 2, S. 109–142.
- McFadden, D. (1978): Modelling the choice of residential location. In: A. Karlqvist; L. Lundqvist; F. Snickars; J. Weibull (Hrsg.): Spatial interaction theory and planning models, Amsterdam: North Holland, S. 75–96.
- Raab, S.; Meisdrock, C. (2015): Bereitstellung der Daten des Zensus 2011 über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder, FDZ-Arbeitspapier Nr. 48. URL: http://www.forschungsdatenzentrum.de/publikationen/veroeffentlichungen/fdz_arbeitspapier-48.pdf, Stand: 23.06.2017.
- Statistische Ämter des Bundes und der Länder (2015): Zensus 2011 – Methoden und Verfahren. Wiesbaden. URL: https://www.zensus2011.de/SharedDocs/Downloads/DE/Publikationen/Aufsaeetze_Archiv/2015_06_MethodenUndVerfahren.pdf?__blob=publicationFile&v=6, Stand: 25.04.2017.
- Tag, K.; Voy, K. (2012): Der Zensus 2011 für Berlin und Brandenburg – Methoden, Datenangebot, Fortschreibungen. Zeitschrift für amtliche Statistik Berlin Brandenburg, Heft 4/2012, S. 54–58.